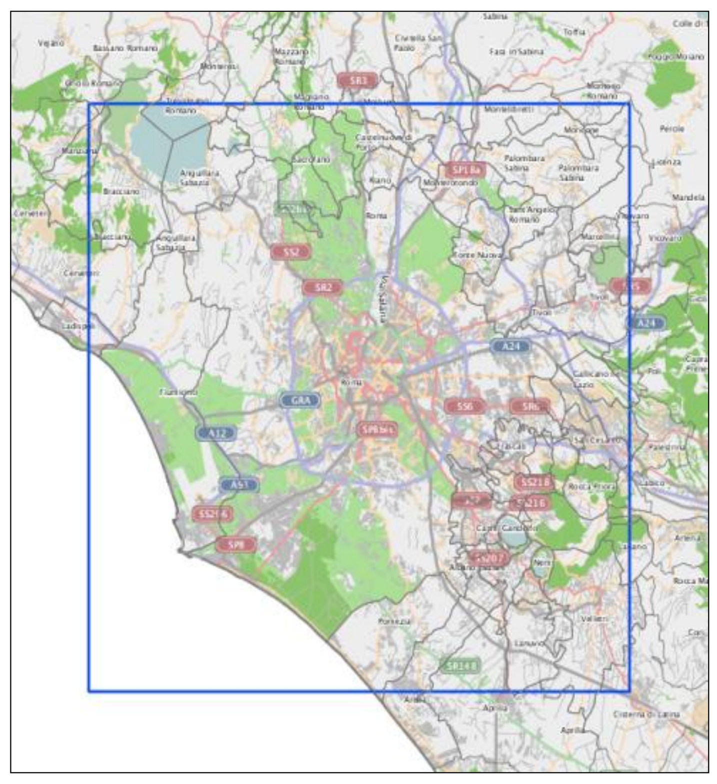
PAHS URBAN CONCENTRATIONS MAPS USING SUPPORT VECTOR MACHINE

Armando Pelliccioni, Andrea Cristofari, Mafalda Lamberti, Claudio Gariazzo
Inail *Research, Via Fontana Candida 1, 00040 Monte Porzio Catone (RM), Italy*

The data



- An area 60 km × 60 km around the urban area of Rome has been considered in one year period (June, 2011 – May, 2012).
- 184 PAHs actual measurements inferred by monitoring stations were available and derived from some 2-10 days field campaigns distributed over all the seasons.
- Data set contained daily samples of three kinds of variables:
  a) meteorological variables (wind direction, wind speed, pressure, precipitations, relative humidity, temperature and total cloud cover)
  b) emissions variables
  c) output produced by a Chemical Transport Model (FARM)
- Also the dates (day and month) have been considered as input variables.

The goal was to build a model able to reproduce PAHs concentrations for constructing exposure maps

The Support Vector Machines (SVM) model

Machine Learning (ML) methods have been applied for forecasting PAHs concentrations.
The purpose of ML methods is to *learn* the rules (supposed to be unknown) governing a system from a set of samples. The learning phase consists in an adaptive process which provides an analytic output function. Support Vector Machines (SVMs) are one of the most widely used ML methods.
In this work, a class of SVM methods, call ε-SVR, has been applied using LIBSM software.
In order to build a model which reproduces well the PAHs concentrations, two main issues have been faced:

1. choosing the samples that represent better the urban dispersion
2. feature selection and normalization

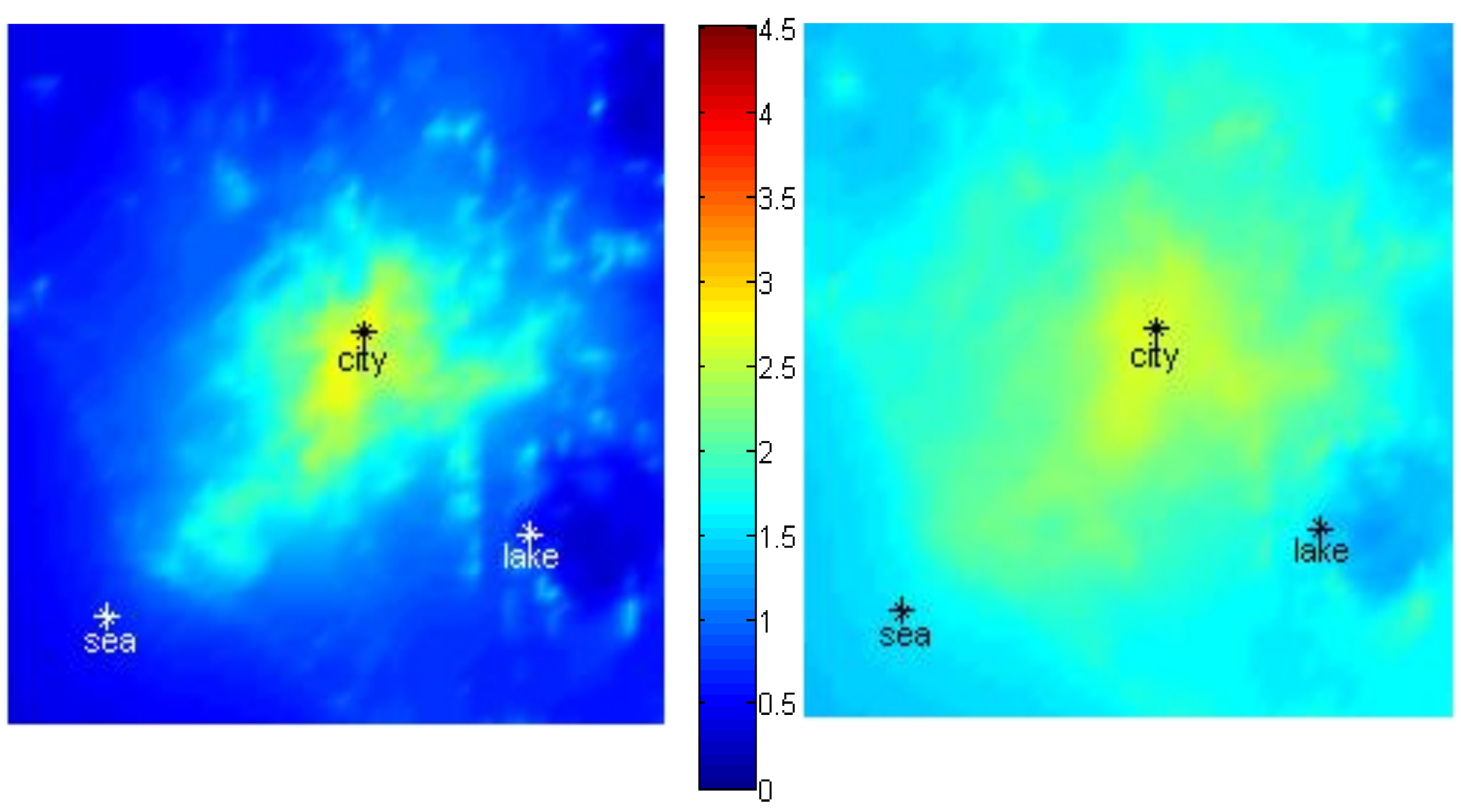
As for the first problem, 124 samples from 16 stations have been selected as training set, and the remaining 60 stations (corresponding to 10 stations) have been used as test. All the training stations are located within the urban area, while some of the testing stations are outside. In this way, test results are expected to provide a significant assessment of the model behavior when it is applied to samples representing different conditions from those of the training samples.
As for the second problem, a preliminary study of the variables distributions has been conducted for choosing an appropriate normalization. After a feature selection process, the following input variables have been finally selected:

- date
- wind direction
- wind speed
- precipitations
- total cloud cover
- output by FARM bc

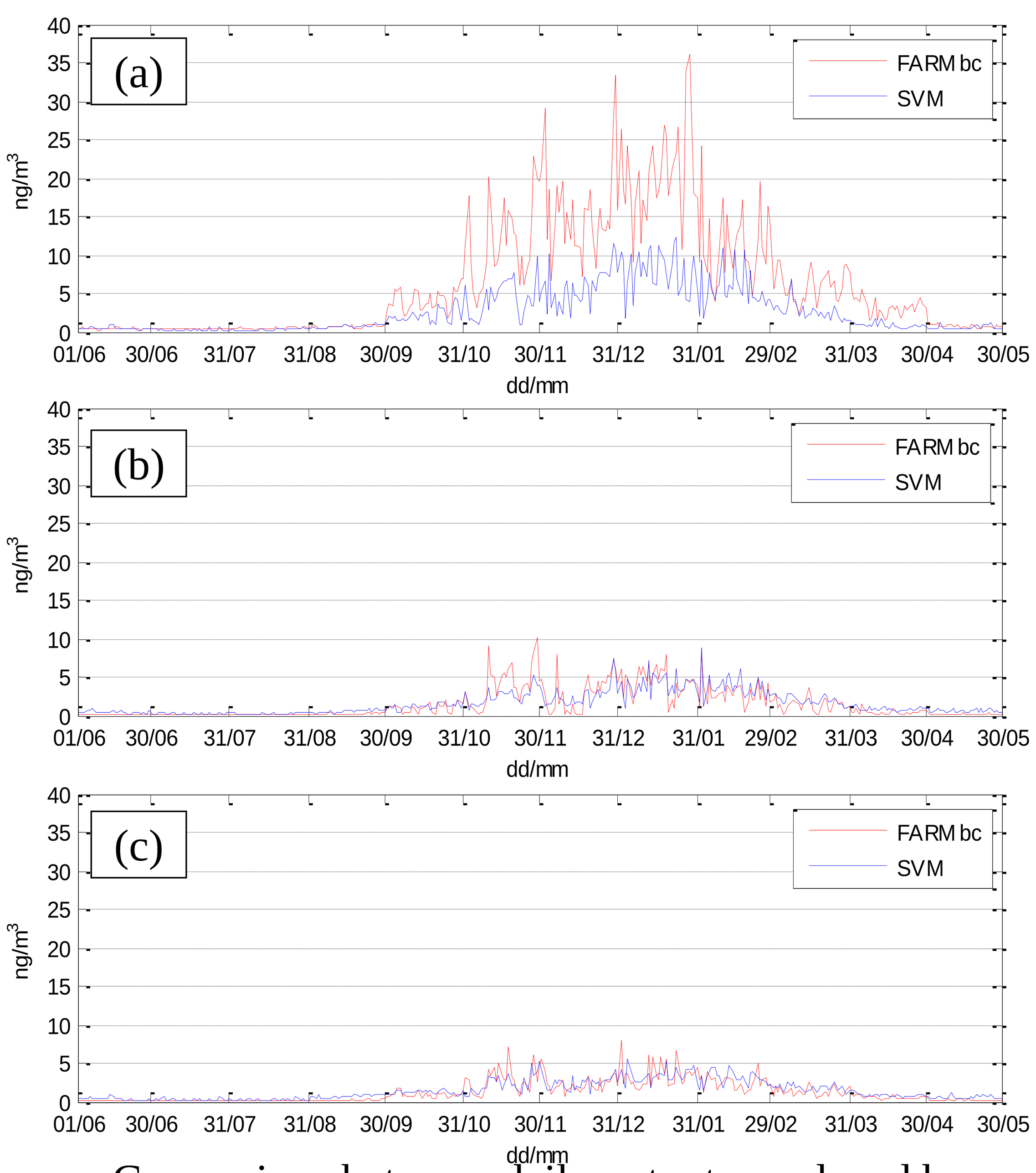
Then, the model has been applied to all the daily samples to construct daily exposure maps.
Since not all the actual measurements were available, two indirect performances indices have been introduced to assess the maps: R_{neg} measures the percentage of negative concentrations, and R_{U-NU} measuares the percentage when urban concentrations are lower than non-urban concentrations.

Results

Annual mean PAHs exposure map by Farm bc | Annual mean PAHs exposure map by SVM



ng/m^3



Comparison between daily outputs produced by SVM and by FARM bc in three different sites: city (a), sea (b), lake (c).

Although the SVM has been trained with only urban samples, it provides congruent estimates (higher in urban area than in non-urban area)

$R_{neg} = 0$

R_{U-NU} = 2,74% – 3,29%

Conclusions

SVM methods have been applied to forecast PAHs concentrations starting from some actual measurements. After a feature selection phase, the test results show a high correlation and performances that improve those achieved by FARM. The model has been applied to construct daily PAHs exposure maps, producing congruent estimates.

With the contribution of the LIFE financial instrument of the European Community